Modeling Home Sale Prices in Ames, Iowa

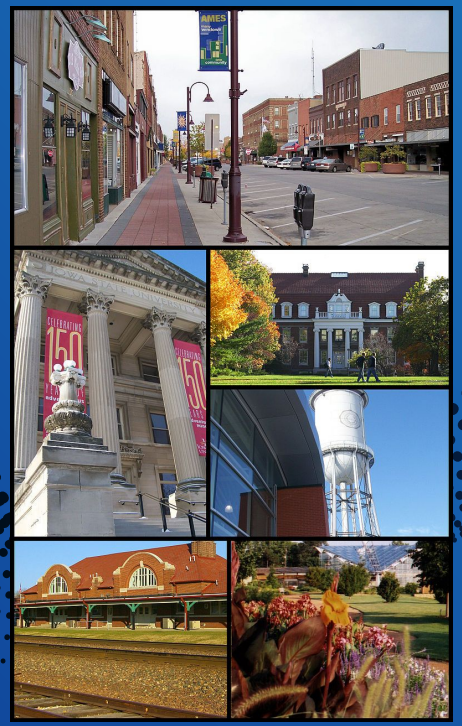
Prepared for: Real Estate Agents, Home Appraisers, Home Buyers/Owners

Brian Cho, Data Scientist

Agenda

- Project Overview
- 2006-2010 house sales
- Predicting the sale price
- Evaluating the model
- Conclusion and Recommendations

Project Overview



Location: Ames, Iowa

Population: ~60,000 (2010)

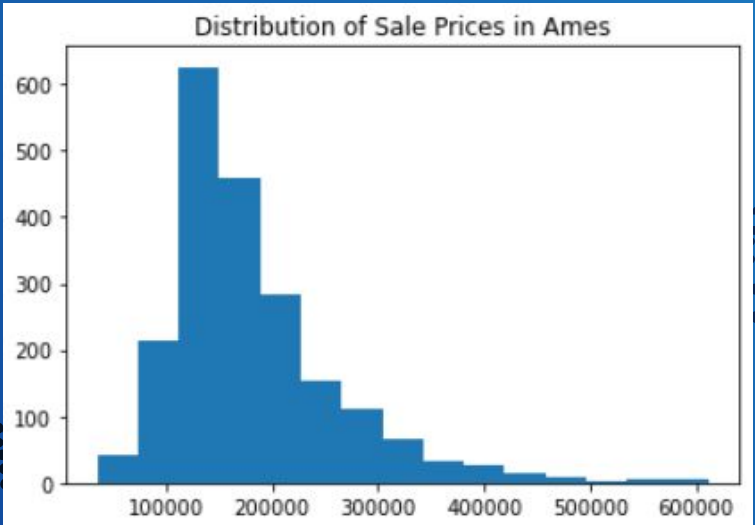
Equip real estate agents and home owners/buyers the ability to accurately assess the value of their property on their own

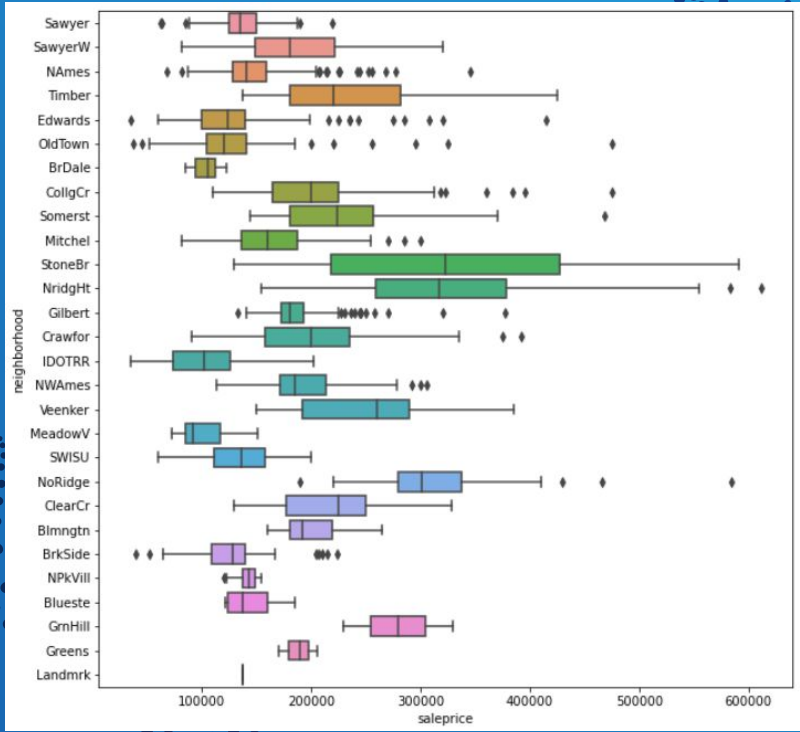
1. Create a model that serves real estate agents, home buyers/sellers in Ames, Iowa
2. Success determined by the linear regression metrics (R2, MSE, RMSE)

2006 - 2010 House Sales

Mean: $181469
Std: $79,258

n = 2051 houses





Predicting the Sale Price

1. Filter and clean 80 variables
2. Ensure proper assumptions are made for a linear regression model
3. Create an initial model - compare to the baseline
4. Refine the model - feature selection, Ridge, LASSO regression
5. Evaluate

Model Metrics

METRIC	OLS	RIDGE	LASSO
R2	Train: 92% Test: 86%	Train: 92% Test: 87%	Train: 91% Test: 87%
RMSE	Train: 19867 Test: 26369	Train: 19866 Test: 25848	Train: 21159 Test: 25373

**Baseline model of taking the average sale price of homes had a RMSE of 80,000*

Conclusion and Recommendations

The Ridge model:

- Best bias-variance tradeoff
- Average error of $22000 in predicting the sale price
- Serves as a better model than seeing the average price of homes in Ames

Recommendations:

- Reduce Complexity in the Model
- Incorporate more data